

Functional Central Limit Theorems for Stick-Breaking Priors*

Yaozhong Hu[†] and Junxi Zhang[‡]

Abstract. We obtain the strong law of large numbers, Glivenko-Cantelli theorem, central limit theorem, functional central limit theorem for various Bayesian nonparametric priors which include the stick-breaking process with general stick-breaking weights, the two-parameter Poisson-Dirichlet process, the normalized inverse Gaussian process, the normalized generalized gamma process, and the generalized Dirichlet process. For the stick-breaking process with general stick-breaking weights, we introduce two general conditions such that the central limit theorem and functional central limit theorem hold. Except in the case of the generalized Dirichlet process, since the finite dimensional distributions of these processes are either hard to obtain or are complicated to use even they are available, we use the method of moments to obtain the convergence results. For the generalized Dirichlet process we use its marginal distributions to obtain the asymptotics although the computations are highly technical.

Keywords: Bayesian Nonparametric priors, stick-breaking process, Dirichlet process, two-parameter Poisson-Dirichlet process, normalized inverse Gaussian process, normalized generalized Gamma process, generalized Dirichlet process, strong law of large numbers, functional central limit theorem.

MSC2020 subject classifications: Primary 60G57, 62F15; secondary 62G20, 60F05.

1 Introduction

Ever since the work of Ferguson (1973) the Dirichlet process has become a critical tool in Bayesian nonparametric statistics and has found applications in various areas, including machine learning, biological science, social science and so on. One of the important features of the Dirichlet process is that when the prior is a Dirichlet process its posterior is also a Dirichlet process (see e.g. Ferguson, 1973). This makes the complex computation in the Bayesian nonparametric analysis possible and enables the Dirichlet process to become a backbone of the Bayesian nonparametric statistics.

To widen the applicability of the Bayesian nonparametric statistics, researchers have tried to extend the concept of Dirichlet process. One of these efforts is the introduction of the stick-breaking process. The first breakthrough along this path is due to

*Yaozhong Hu was supported by an NSERC discovery fund and a startup fund of University of Alberta.

[†]Department of Mathematical and Statistical Sciences, University of Alberta at Edmonton, Alberta, Canada, T6G 2G1, yaozhong@ualberta.ca

[‡]Department of Mathematical and Statistical Sciences, University of Alberta at Edmonton, Alberta, Canada, T6G 2G1, junxi3@ualberta.ca

Sethuraman (1994) who shows that the Dirichlet process admits the stick-breaking representation (see (2.1)–(2.2) in the next section), where the stick-breaking weights are independent and identically distributed (i.i.d.) random variables satisfying the Beta distribution $\text{Beta}(1, a)$ (throughout this paper the notation $\text{Beta}(\alpha, \beta)$ denotes the Beta distribution whose density is $g(x; \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$). Within this stick-breaking representation, we can extend the class of Dirichlet processes to many other priors by assuming that the stick-breaking weights are i.i.d. with other distributions; satisfy some other kinds of dependence; or satisfy some specific (joint) distributions. Among various such extensions, let us mention the following works which we shall deal with in this paper. Perman et al. (1992) obtain a general formulae for sized-biased sampling from a Poisson point process where the size of a point is defined by an arbitrary strictly positive function. From these formulae, they identify the stick-breaking representation of the two-parameter Poisson-Dirichlet process, which admits a stick-breaking process with the stick-breaking weights $v_i \stackrel{\text{ind}}{\sim} \text{Beta}(1-b, a+ib)$, where $b > 0$, $a > -b$ and $i = 1, 2, \dots$. Favaro et al. (2012) introduce the normalized inverse Gaussian process through its stick-breaking representation by identifying the explicit finite dimensional joint density functions of its stick-breaking weights. Favaro et al. (2016) present the stick-breaking representation of homogeneous normalized random measures with independent increments (hNRMIs) (see e.g. Regazzini et al., 2003 for more details of NRMIs), which include the normalized generalized gamma process and the generalized Dirichlet process, two widely used priors in Bayesian nonparametric statistics.

Strong law of large numbers, central limit theorem and functional central limit theorem have always been ones of the central topics in statistics and in probability theory. Without exception the asymptotic behaviors of the Dirichlet process and other Bayesian nonparametric priors play important roles in the Bayesian nonparametric analysis, for example in the construction of asymptotic Bayesian confidence intervals, regression analysis and functional estimations. Compared to the vast literature in the field of parametric statistics relevant to these issues the achievements in the field of Bayesian nonparametrics are quite limited. However, let us mention the following works pioneered this paper. Sethuraman and Tiwari (1982) discuss the weak convergences of the Dirichlet measure P when its parameter measure (i.e the measure aH in this paper) approaches to a non-zero measure or a zero measure respectively. Lo (1983) studies the central limit theorem of the posterior distribution of Dirichlet process which is analogous to our central limit theorem for the Dirichlet process. Based on this result, Lo (1987) obtains the asymptotic confidence bounds and establishes the asymptotic validity of the Bayesian bootstrap method. The above mentioned Lo's results are extended to the mixtures of Dirichlet process by Brunner and Lo (1996). James (2008) reveals the consistency behavior (the posterior distribution converges to the true distribution weakly) and the functional central limit theorem for the posterior distribution of the two-parameter Poisson-Dirichlet process (with fixed a and when the sample size goes to infinity). The consistency of the posterior is discussed by Ho Jang et al. (2010) when the priors are the two-parameter Poisson-Dirichlet prior and the species sampling prior. Furthermore, De Blasi et al. (2013) investigate the consistency of the Gibbs-type priors. Kim and Lee (2004) show that the Bernstein-von Mises theorem holds in survival

models for the Dirichlet process, Beta process and Gamma process. Dawson and Feng (2006) establish the large deviation principle for the Poisson-Dirichlet distribution and give the explicit rate functions when the parameter a (which represents the mutation rate in the context of population genetics) approaches infinity. Labadi and Zarepour (2013) present the functional central limit theorem for the normalized inverse Gaussian process on $D(\mathbb{R})$ when its parameter a is large by using its finite dimensional joint density. Labadi and Abdelrazeq (2016) obtain the functional central limit theorem for the Dirichlet process by using the finite dimensional densities and for the Beta process on $D(\mathbb{R})$ by using the characteristic function.

From the above mentioned works we see that there are only very limited results on the asymptotics of the stick-breaking processes. Relevant to the asymptotics as $a \rightarrow \infty$, there have been established the central limit theorem and functional central limit theorem only for two processes: the Dirichlet process and the normalized inverse Gaussian process. The reason for the above limitation is that the most commonly used technique appeals to the explicit forms of the finite dimensional densities of the process itself. This method is effective only when the finite dimensional distributions have explicit forms and are possible to handle. It cannot be applied to study other processes when the explicit forms for the finite dimensional marginal densities of the process itself are unavailable or they are too complex to analyze even though they are available.

This paper is to introduce the method of moments into this study and to provide a systematic study of the asymptotics as $a \rightarrow \infty$ for various stick-breaking processes depending on a parameter $a > 0$. Let us emphasize that the method of moments in this paper refers to the fact that if the distribution of the random variable X is determined by its moments, and the random variables $\{X_i\}_{i=1}^n$ have all moments, and if $\lim_{n \rightarrow \infty} \mathbb{E}[X_n^r] = \mathbb{E}[X^r]$ for $r = 1, 2, \dots$, then $X_n \xrightarrow{d} X$ (see e.g. Billingsley, 1995, Theorem 30.2). We are mainly concerned with three types of the asymptotics (strong law of large numbers, central limit theorem, and functional central limit theorem) for a number of processes, which include the stick-breaking process with general stick-breaking weights, the classical Dirichlet process $DP(a, H)$ (see Ferguson, 1973), the two-parameter Poisson-Dirichlet process $PDP(a, b, H)$ (also known as Pitman-Yor process, Pitman and Yor, 1997), the normalized inverse Gaussian process $N-IG(a, H)$ (see Lijoi et al., 2005b), the normalized generalized gamma process $NGG(\sigma, a, H)$ (see Lijoi et al., 2003, 2007; Brix, 1999), and the generalized Dirichlet process $GDP(a, r, H)$ (see Lijoi et al., 2005a).

All of the mentioned processes depend on a parameter a which is usually called the *concentration parameter*. It is of the same order as the inverse of the variance of the process (see Remark 3.6 for more precise meaning). It has also some more specific meanings for various processes. For example, if $\{X_i\}_{i=1}^n$ is a sample from the Dirichlet process $DP(a, H)$, then it is known that the posterior mean is $\mathbb{E}[P(\cdot)|X_1, \dots, X_n] = \frac{a}{a+n}H(\cdot) + \frac{n}{a+n} \frac{\sum_{i=1}^n X_i}{n}$, which means that a plays the key role of the weight of the prior.

For the generalized Dirichlet process since the finite dimensional marginal distributions of the process itself are available we shall use them to obtain the asymptotics directly although the computations are very technical. Let us point out that this process also admits a stick-breaking representation. However, it seems to us that it is more

complex to use the method of moments than to use the finite dimensional marginal distributions of the process itself.

Let us stress the following points of the paper about the well-known Bayesian non-parametric priors.

- (1) (for Dirichlet process) Both the finite dimensional distributions of the stick-breaking weights and the process itself are explicit and are easy to handle. Prior to this work the central limit theorem and the functional central limit theorem have been established for this process by using the finite dimensional distribution of the process itself.

For the Dirichlet process the stick-breaking weights $\{v_i\}$ are i.i.d. and follow the Beta distribution $\text{Beta}(1, \alpha)$. We introduce the concept of stick-breaking process with general stick-breaking weights, where we still require the stick-breaking weights $\{v_i\}$ to be i.i.d. but the law μ they follow can be arbitrary. In this case there is no way to obtain the explicit form of the joint distributions of the process itself. We use the method of moments to establish the central limit theorem and the functional central limit theorem for this process. For example, $v_i \sim \text{Beta}(\rho_a, a)$, where ρ_a is a function of a such that $\rho_a/a \rightarrow 0$ as $a \rightarrow \infty$. In this case the joint distributions of the process itself are unavailable except in the case $\rho_a = 1$, i.e. in the case of the Dirichlet process.

- (ii) (for the normalized inverse Gaussian process and for the generalized Dirichlet process) Both the finite dimensional distributions of the stick-breaking weights and that of the process itself are explicit. Prior to this work the central limit theorem and the functional central limit theorem have been established only for the normalized inverse Gaussian process by using the finite dimensional distributions of the process itself. We shall also use the finite dimensional distributions of the process itself to obtain the central limit theorem and the functional central limit theorem for the generalized Dirichlet process. We shall use the method of moments to re-derive the central limit theorem and the functional central limit theorem for the normalized inverse Gaussian process, providing an alternative tool for this process.
- (iii) (for the two-parameter Poisson-Dirichlet process and the normalized generalized gamma process) The finite dimensional distributions of the stick-breaking weights are known but the finite dimensional distributions of the process itself are not available. We use the method of moments to obtain the central limit theorem and the functional central limit theorem for these processes.

Now we explain the organization of this paper. In Section 2, we recall the general stick-breaking process and introduce the stick-breaking process with general stick-breaking weights ($\text{SPG}(\mu, H)$). In Section 3, we present the moment results for various stick-breaking processes, including $\text{SPG}(\mu, H)$, $\text{PDP}(a, b, H)$, $\text{N-IG}(a, H)$, and $\text{NGG}(\sigma, a, H)$, $\text{GDP}(a, r, H)$ separately since the computations are different for different processes. In Section 4, we state the strong law of large numbers, central limit theorem,

and functional central limit theorem. The stick-breaking process with general stick-breaking weights are new and we allow the stick-breaking weights to be some very general i.i.d. random variables defined on $(0, 1)$. With different choices of the stick-breaking weights we can obtain various known stick-breaking processes. Because of this generality of the stick-breaking weights we state one theorem on the central limit theorem and functional central limit theorem for this type of processes. We state a similar theorem for all other processes (PDP(a, b, H), N-IG(a, H), NGG(σ, a, H), GDP(a, r, H)). The details of the proofs will be provided in a supplementary file (Hu and Zhang, 2021) where we also include some definitions and some well-known propositions of the mentioned processes to provide the necessary background. Interested readers are referred to Hu and Zhang (2020) and references therein for a recent survey of some of these processes and their applications.

Finally, let us emphasize that all the processes we dealt with in this paper are actually “random probability measures”. However, we follow the convention in the literature to continue to call them “processes”.

2 Preliminary notations

2.1 Definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and let $(\mathbb{X}, \mathcal{X})$ be a measurable Polish space, namely, \mathbb{X} is a separable complete metric space and \mathcal{X} is the Borel σ -algebra of \mathbb{X} . Let H be a nonatomic probability measure on $(\mathbb{X}, \mathcal{X})$ (i.e. $H(\{x\}) = 0$ for any $x \in \mathbb{X}$). Now we give the definition of the stick-breaking process (more appropriately a stick-breaking random probability measure).

Definition 2.1. A random measure $P = (P(\omega, A), \omega \in \Omega, A \in \mathcal{X})$ is said to be a stick-breaking process with the base measure H , if it has the following representation:

$$\left\{ \begin{aligned} P &= \sum_{i=1}^{\infty} w_i \delta_{\theta_i}, & \text{where} & & (2.1) \\ w_1 &= v_1, & w_i &= v_i \prod_{j=1}^{i-1} (1 - v_j) & \text{for } i = 2, 3, \dots, & (2.2) \end{aligned} \right.$$

where $\theta_i, i = 1, 2, \dots$ are i.i.d. random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(\mathbb{X}, \mathcal{X})$ such that for each i , the law of θ_i is H ; δ_{θ_i} denotes the Dirac measure on $(\mathbb{X}, \mathcal{X})$, and $v_i, i = 1, 2, \dots$ are random variables with values in $[0, 1]$, independent of $\{\theta_i\}$, which are called the stick-breaking weights.

Since we assume that $\{\theta_i\}$ are i.i.d. and follow the distribution H , if H is given and fixed, then the random probability measure P depends only on the choice of $\{v_i\}$.

Remark 2.2. To make sure that P is well-defined (namely, (2.1) is convergent), one needs to impose the condition that $\sum_{i=1}^{\infty} w_i = 1$ almost surely, which is equivalent to the condition that $\sum_{i=1}^{\infty} \log \mathbb{E}[(1 - v_i)] = -\infty$ (e.g. Ghosal and Van der Vaart, 2017, Lemma 3.4).

Remark 2.3. Throughout the entire paper, we shall assume that a is a positive real number and H is a nonatomic measure on $(\mathbb{X}, \mathcal{X})$ unless otherwise specified.

For potential applications in practice we introduce the concept of *stick-breaking process with general stick-breaking weights*.

Definition 2.4. P is called the *stick-breaking process with general stick-breaking weights*, denoted by $P \sim \text{SPG}(\mu, H)$, if the stick-breaking weights $\{v_1, v_2, \dots\}$ in (2.1)–(2.2) are i.i.d. and follow a general distribution μ .

Remark 2.5. The law μ on $(0, 1)$ can be of continuous or discrete types, or the mixture. An interesting special example is the quasi Bernoulli stick-breaking process (Zeng and Duan, 2020), where the $v_i \sim g(x) = pf(x) + \frac{1-p}{\varepsilon}f(x/\varepsilon)$ for the Bernoulli density $f(x) \sim \text{Beta}(1, a)$ and for some $p \in (0, 1), \varepsilon > 0$.

Based on the expectation and variance of P , we introduce the following quantities that are investigated in the main theorems:

$$D_a(\cdot) = \frac{P(\cdot) - \mathbb{E}[P(\cdot)]}{\sqrt{\text{Var}[P(\cdot)]}} = \frac{P(\cdot) - H(\cdot)}{\sqrt{H(A)(1 - H(A))\mathbb{E}[\sum_{i=1}^{\infty} w_i^2]}}, \tag{2.3}$$

where the last identity follows from (7.12)–(7.13) (in the supplementary material). Up to a constant we may just consider the following quantity for notational simplicity:

$$Q_{H,a}(\cdot) = \frac{P(\cdot) - \mathbb{E}[P(\cdot)]}{\sqrt{\mathbb{E}[\sum_{i=1}^{\infty} w_i^2]}}. \tag{2.4}$$

3 Moment results

We use the method of moments to show the announced asymptotics. This requires to have some nice estimates of the moments of the random probability measure P , which in turn requires some nice bounds for the moments of $\{w_i\}_{i=1}^{\infty}$. Thus, in this section we present the asymptotic behaviors of the joint moments of w_i 's for various processes introduced in the introduction. These results will play the key roles in the proofs of our main theorems. On the other hand, they also have their own interest.

In the following proposition and throughout the paper we use the notation $p_{m:n} := \sum_{i=m}^n p_i$ for $m \leq n$, and let the sequence $\{w_i\}_{i=1}^{\infty}$ be defined as in (2.2).

Proposition 3.1. Let $P \sim \text{SPG}(\mu_a, H)$, i.e., the law of the i.i.d stick-breaking weights v_i is μ_a , where $a > 0$ is a certain parameter. We assume that v_i is not identically 0. If $\lim_{a \rightarrow \infty} \frac{\mathbb{E}[v_1^{n+1}]}{\mathbb{E}[v_1^n]} = 0$ for all $n \in \mathbb{Z}_+$ (set of nonnegative integers), then for any nonnegative integers m, n ,

$$\mathbb{E}[v_i^n (1 - v_i)^m] = \mathbb{E}[v_1^n] + o(\mathbb{E}[v_1^n]), \tag{3.1}$$

$$\sum_{j=0}^{\infty} (\mathbb{E}[(1 - v_i)^m])^j = \frac{1}{m\mathbb{E}[v_1]} + o\left(\frac{1}{m\mathbb{E}[v_1]}\right). \tag{3.2}$$

Furthermore, for any positive integers p_1, \dots, p_k , we have

$$\begin{aligned} & \mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \dots w_{i_k}^{p_k} \right] \\ &= \frac{\mathbb{E}[v_1^{p_1}] \dots \mathbb{E}[v_1^{p_k}]}{p_{1:k} p_{2:k} \dots p_{k:k} (\mathbb{E}[v_1])^k} + o \left(\frac{\mathbb{E}[v_1^{p_1}] \dots \mathbb{E}[v_1^{p_k}]}{(\mathbb{E}[v_1])^k} \right). \end{aligned} \tag{3.3}$$

In particular, when $p_j = 2$ for all $j \in \{1, \dots, k\}$ (hence $p_{1:k} = 2k$), the asymptotics (3.3) becomes

$$\mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^2 w_{i_2}^2 \dots w_{i_k}^2 \right] = \frac{1}{2^k k!} \left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]} \right)^k + o \left(\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]} \right)^k \right). \tag{3.4}$$

Proposition 3.2. Let $P \sim \text{PDP}(a, b, H)$. Namely, let the stick-breaking weights v_1, v_2, \dots be given by (7.2) (in the supplementary material). Then, for any positive integers p_1, \dots, p_k , we have the following identity.

$$\begin{aligned} & \mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \dots w_{i_k}^{p_k} \right] \\ &= \frac{1}{(a + kb)(a + 1)_{(p_{1:k}-1)}} \prod_{i=1}^k \frac{(1 - b)_{p_i} (a + bi)}{p_{i:k} - (k - i + 1)b}. \end{aligned} \tag{3.5}$$

In particular, when $p_j = 2$ for all $j \in \{1, \dots, k\}$, the above expectation becomes

$$\mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^2 w_{i_2}^2 \dots w_{i_k}^2 \right] = \frac{(1 - b)^k (a + b) \dots (a + b(k - 1))}{k! (a + 1) \dots (a + 2k - 1)}. \tag{3.6}$$

Proposition 3.3. Let $P \sim \text{N-IG}(a, H)$. Namely, let the stick-breaking weights $\{v_i\}_{i=1}^\infty$ be given by (7.3)–(7.4) (in the supplementary material). Then, for any positive integers p, p_1, \dots, p_k , we have

$$\mathbb{E} \left[\sum_{n=1}^\infty w_n^p \right] = O \left(\frac{1}{a^{p-1}} \right) \quad \text{as } a \rightarrow \infty, \tag{3.7}$$

$$\mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \dots w_{i_k}^{p_k} \right] = O \left(\frac{1}{a^{p_{1:k} - k}} \right) \quad \text{as } a \rightarrow \infty. \tag{3.8}$$

Furthermore, when $p = p_1 = \dots = p_k = 2$, we have

$$\mathbb{E} \left[\sum_{n=1}^\infty w_n^2 \right] = \frac{1}{a} + o \left(\frac{1}{a} \right) \quad \text{as } n \rightarrow \infty, \tag{3.9}$$

$$\mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^2 w_{i_2}^2 \dots w_{i_k}^2 \right] = \frac{1}{k!a^k} + o\left(\frac{1}{a^k}\right). \tag{3.10}$$

[Namely, the leading coefficient in (3.7) is 1 and the leading coefficient in (3.8) is $\frac{1}{k!}$.]

Proposition 3.4. Let $P \sim \text{NGG}(\sigma, a, H)$. Namely, let the distribution of the stick-breaking weights $\{v_1, v_2, \dots\}$ be given by (7.5)–(7.6) (in the supplementary material). Then, for any positive integers p_1, \dots, p_k , we have

$$\mathbb{E} \left[\sum_{n=1}^{\infty} w_n^{p_1} \right] = O\left(\frac{1}{a^{p_1-1}}\right) \quad \text{as } a \rightarrow \infty, \tag{3.11}$$

$$\mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \dots w_{i_k}^{p_k} \right] = O\left(\frac{1}{a^{p_1+k-k}}\right) \quad \text{as } a \rightarrow \infty. \tag{3.12}$$

Furthermore, when $p = p_1 = \dots = p_k = 2$ and when $\sigma = \frac{1}{m}$ for some arbitrarily fixed integer $m \geq 2$, we have

$$\mathbb{E} \left[\sum_{n=1}^{\infty} w_n^2 \right] = \frac{1}{a} + o\left(\frac{1}{a}\right) \quad \text{as } a \rightarrow \infty, \tag{3.13}$$

$$\mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^2 w_{i_2}^2 \dots w_{i_k}^2 \right] = \frac{1}{k!a^k} + o\left(\frac{1}{a^k}\right) \quad \text{as } a \rightarrow \infty. \tag{3.14}$$

Proposition 3.5. Let $P \sim \text{GDP}(a, r, H)$ and let p_1, \dots, p_k be positive integers. Then, as $a \rightarrow \infty$,

$$\mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \dots w_{i_k}^{p_k} \right] = O\left(\frac{1}{a^{p_1+k-k}}\right). \tag{3.15}$$

In particular, when $p_j = 2$ for all $j \in \{1, \dots, k\}$, the above expectation becomes

$$\mathbb{E} \left[\sum_{1 \leq i_1 < i_2 < \dots < i_k < \infty} w_{i_1}^2 w_{i_2}^2 \dots w_{i_k}^2 \right] = \frac{\left[\frac{\sum_{k=1}^r \left(\frac{1}{k}\right)^2}{\left(\sum_{j=1}^r \frac{1}{j}\right)^2} \right]^k}{k!a^k} + o\left(\frac{1}{a^k}\right). \tag{3.16}$$

Remark 3.6. As for $\text{SPG}(\mu_a, H)$, a is a parameter such that v_i converges in distribution to 1 as $a \rightarrow \infty$. And we will give more details in Remark 4.13 on page 13 later on. For the specified processes in Propositions 3.2–3.5, the parameter a is the prior precision or the concentration parameter as we mentioned in the introduction. We can also see that the parameter a is the same order as $\frac{1}{\mathbb{E}[\sum_{i=1}^{\infty} w_i^2]}$.

Remark 3.7. The special cases when $p_1 = \dots = p_k = 2$ in Propositions 3.1–3.5 are particularly important, since the corresponding terms in Theorem 4.4 will not converge

to zero and we need to use them to identify the limits. Other terms will converge to 0. This is because otherwise some p_i will be greater than 2 and then there will be fewer factors in the product for the same value $p_1 + \dots + p_k$ (see the proof of Theorem 4.4, Cases 1 and 2).

Remark 3.8. The quantity $p(n_1, \dots, n_k) = \sum_{i_1, \dots, i_k} \mathbb{E} [w_{i_1}^{n_1} w_{i_2}^{n_2} \dots w_{i_k}^{n_k}]$ bears the same form of the exchangeable partition probability function (EPPF) in the random partition theory (see e.g Pitman, 1996, 2003), where the w_i is replaced by the so-called size biased permutation from a random partition. In the study of Poisson-Kingman model, the order statistics w_1^*, w_2^*, \dots of w_1, w_2, \dots are given by $w_i^* = \frac{J_i}{J_1 + J_2 + \dots}$, where J_1, J_2, \dots are the ranked points of a Poisson process with Lévy density ρ (see Pitman, 2003, Definition 3). When v_i 's are iid Beta(1, a), w_i^* is Poisson-Dirichlet distribution (see Pitman, 1996, Theorem 5). In general case it seems hard to find the distribution of w_i^* from v_i 's. However, it remains interesting to apply our method of moments to study the asymptotics for the Poisson-Kingman model.

4 Main results

4.1 Strong law of large numbers

The strong law of large numbers and the Glivenko-Cantelli theorem play undoubtedly important roles in statistics. In this subsection we state the strong law of large numbers and the Glivenko-Cantelli theorem for various processes introduced in the introduction. But before we state our theorem, we need an additional condition on the stick-breaking weights v_i in the case of SPG(μ_a, H).

Assumption 4.1. Let the i.i.d. stick-breaking weights $\{v_i\}$ satisfy

$$\mathbb{E}[v_i^p] = \frac{C_p}{a^{k_p}} + o\left(\frac{1}{a^{k_p}}\right) \quad \text{as } a \rightarrow \infty, \tag{4.1}$$

for any $p \in \mathbb{N}$, where k_p is a positive sequence satisfying $jk_i \geq ik_j$ for $i \geq j$ and C_p is a sequence of finite constants, independent of a .

Theorem 4.2. Let P be one of the stick-breaking process with general stick-breaking weights SPG(μ_a, H) satisfying Assumption 4.1, the two-parameter Poisson-Dirichlet process PDP(a, b, H), the normalized inverse Gaussian process N-IG(a, H), the normalized generalized gamma process NGG(σ, a, H), and the generalized Dirichlet process GDP(a, r, H). Assume that $a = N^\tau$ for some arbitrarily fixed $\tau > 0$. Then, as $N \rightarrow \infty$,

$$P(A) \xrightarrow{a.s.} H(A) \tag{4.2}$$

for any measurable set $A \in \mathcal{X}$.

Once we have the strong law of large numbers for P , we can deduce the Glivenko-Cantelli theorem for P (see e.g. Theorem 20.6 in Patrick (1995) for a general discussion).

Theorem 4.3. Let $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Let P be one of the stick-breaking process with general stick-breaking weights $\text{SPG}(\mu_a, H)$ satisfying Assumption 4.1, the two-parameter Poisson-Dirichlet process $\text{PDP}(a, b, H)$, the normalized inverse Gaussian process $\text{N-IG}(a, H)$, the normalized generalized gamma process $\text{NGG}(\sigma, a, H)$, and the generalized Dirichlet process $\text{GDP}(a, r, H)$. Assume that $a = N^\tau$ for some arbitrarily fixed $\tau > 0$. Then, as $N \rightarrow \infty$,

$$\sup_{x \in \mathbb{R}} |P((-\infty, x]) - H((-\infty, x])| \xrightarrow{a.s.} 0.$$

4.2 Central limit theorems and functional central limit theorems

In this subsection, we state the central limit theorems corresponding to the strong law of large numbers of the form (4.2).

We shall state the central limit theorems and functional central limit theorems for various processes as the following three theorems. The first one is for the stick-breaking process with general stick-breaking weights defined by Definition 2.4. We will assume mild convergence conditions on the stick-breaking weights.

Theorem 4.4. Let $P \sim \text{SPG}(\mu_a, H)$, where the stick-breaking weights v_1, v_2, \dots (whose distributions) depending on a parameter $a > 0$ (we omit the explicit dependence on a of the v_i 's). Let D_a and $Q_{H,a}$ be defined by (2.3) and (2.4) respectively. Assume that the stick-breaking weights v_1, v_2, \dots satisfy the following two conditions.

(i) For all $n \in \mathbb{Z}^+$, we have

$$\lim_{a \rightarrow \infty} \frac{\mathbb{E}[v_1^{n+1}]}{\mathbb{E}[v_1^n]} = 0. \tag{4.3}$$

(ii) For any multi-index (p_1, \dots, p_k) such that $p_i \geq 2$ and $\frac{p_{1:k}}{2} > k$, where $p_{1:k} = \sum_{i=1}^k p_i$, we have

$$\lim_{a \rightarrow \infty} \frac{(\mathbb{E}[v_1])^{\frac{p_{1:k}}{2} - k} \prod_{i=1}^k \mathbb{E}[v_1^{p_i}]}{(\mathbb{E}[v_1^2])^{\frac{p_{1:k}}{2}}} = 0. \tag{4.4}$$

Then we have the following results.

(i) (Central limit theorem) Let A_1, A_2, \dots, A_n be any disjoint measurable subsets of \mathbb{X} . Then, as $a \rightarrow \infty$,

$$(D_a(A_1), D_a(A_2), \dots, D_a(A_n)) \xrightarrow{d} (X_1, X_2, \dots, X_n), \tag{4.5}$$

where $(X_1, X_2, \dots, X_n) \sim N(0, \Sigma)$ and $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq n}$ is given by

$$\sigma_{ij} = \begin{cases} 1 & \text{if } i = j, \\ -\sqrt{\frac{H(A_i)H(A_j)}{(1-H(A_i))(1-H(A_j))}} & \text{if } i \neq j. \end{cases} \tag{4.6}$$

(ii) (Functional central limit theorem) Let $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be the d -dimensional Euclidean space with the Borel σ -algebra. Then

$$Q_{H,a} \xrightarrow{\text{weakly}} B_H^o \quad \text{in} \quad D(\mathbb{R}^d) \tag{4.7}$$

with respect to the Skorohod topology.

Remark 4.5. For central limit theorem we use D_a because each component converges to a standard Gaussian. For functional central limit theorem we use $Q_{H,a}$ since it converges to a Brownian bridge with parameter H . We can presumably use D_a (or $Q_{H,a}$) in both (4.6) and (4.7) with a scaling.

The conditions (i) and (ii) in Theorem 4.4 are implied by many other conditions. One of them is given below.

Remark 4.6. Assumption 4.1 implies the conditions (i) and (ii) in Theorem 4.4.

Proof. It is obvious that $\{k_p\}$ is an increasing sequence, and thus the condition (i) of Theorem 4.4 (i.e (4.3)) holds.

For any nonnegative integer m , let \mathfrak{N} be a certain collection of integers j 's such that $\sum_{j \in \mathfrak{N}} j = m$. The condition (ii) in Theorem 4.4 is equivalent to the following statement:

If $j \geq 2$ and $|\mathfrak{N}| < \frac{m}{2}$, then

$$\frac{m(k_2 - k_1)}{2} < \sum_{j \in \mathfrak{N}} k_j - |\mathfrak{N}|k_1. \tag{4.8}$$

Thus, to prove (4.4) it is sufficient to show $\frac{mk_2}{2} < \sum_{j \in \mathfrak{N}} k_j$. This is a simple consequence of $jk_i \geq ik_j$ for $i \geq j$. In fact, taking $i = 2$, we have for all $j \geq 2$, $2k_j \geq jk_2$ holds and thus we have $\sum_{j \in \mathfrak{N}} 2k_j \geq \sum_{j \in \mathfrak{N}} jk_2$, which implies $\frac{mk_2}{2} < \sum_{j \in \mathfrak{N}} k_j$. Hence we have (4.8). \square

The conditions (i) and (ii) in Theorem 4.4 are satisfied by many interesting processes including the Dirichlet process. We give three examples to illustrate the applicability of our above theorem.

Corollary 4.7. Theorem 4.4 holds true when $P \sim \text{DP}(a, H)$.

Proof. It is sufficient to verify the condition (4.1) in Assumption 4.1. Since $v_i \stackrel{iid}{\sim} \text{Beta}(1, a)$, we have for any positive integer p ,

$$\mathbb{E}[v_i^p] = \frac{\Gamma(a+1)\Gamma(p+1)}{\Gamma(1)\Gamma(a+p+1)} = \frac{p!}{(a+1)\cdots(a+p)} = \frac{p!}{a^p} + o\left(\frac{1}{a^p}\right).$$

Hence, $k_p = p$ and $C_p = p!$. Obviously, for $i \geq j$, $jk_i \geq ik_j$ always holds true. \square

Remark 4.8. Since the posterior of the Dirichlet process is still a Dirichlet process, the above result can be applied to the posterior process in the Bayesian nonparametric models when the prior is the Dirichlet process for the following situations: (i) with large sample size and finite parameter a ; (ii) with large parameter a and finite sample size, (iii) with parameter a and sample size both large.

The assumption of the Beta(1, a)-distribution in Corollary 4.7 can be replaced by a general Beta(ρ_a , a), where $\rho_a/a \rightarrow 0$. In fact, in this case, we have

$$\mathbb{E} [v_1^n] = \left(\frac{\rho_a}{a}\right)^n + o\left(\left(\frac{\rho_a}{a}\right)^n\right).$$

It is easy to verify that the conditions (4.3)–(4.4) in Theorem 4.4 are satisfied. Thus we have

Corollary 4.9. Theorem 4.4 holds true when $P \sim \text{SPG}(\mu_a, H)$, where $v_i \stackrel{iid}{\sim} \text{Beta}(\rho_a, a)$ with $\lim_{a \rightarrow \infty} \frac{\rho_a}{a} = 0$.

Remark 4.10. It is not clear yet what is the finite dimensional distribution of stick-breaking process P if the corresponding stick-breaking weights $v_i \stackrel{iid}{\sim} \text{Beta}(\rho_a, a)$.

The next corollary is about the asymptotic behavior of the prior P , when the corresponding stick-breaking weights v_i follow a linear combination of Beta distributions, whose precise meaning is given below.

Definition 4.11. Let s be any positive integer and let $\{r_1, \dots, r_s\}$ and $\{t_1, \dots, t_s\}$ be two sets of positive real numbers such that $\sum_{\ell=1}^s t_\ell = 1$. Let $u_{1,1}, \dots, u_{1,s}, u_{2,1}, \dots, u_{2,s}, \dots$ be independent and let $u_{i,\ell} \sim \text{Beta}(1, a^{r_\ell}), i = 1, 2, \dots, \ell = 1, \dots, s$. Then the random variables

$$v_i = \sum_{\ell=1}^s t_\ell u_{i,\ell}, \quad i = 1, 2, \dots, \tag{4.9}$$

are called linear combinations of Beta random variables.

Corollary 4.12. Theorem 4.4 holds true when P is the stick-breaking process as defined in Definition 2.1, where the weights v_i are the linear combinations of Beta random variables defined by (4.9).

Proof. By the independence of $\{u_{i,\ell}\}_{\ell=1}^s$, we can compute the p -th moment of v_i as follows.

$$\begin{aligned} \mathbb{E}[v_i^p] &= \mathbb{E} \left[\left(\sum_{\ell=1}^s t_\ell u_{i,\ell} \right)^p \right] = \sum_{\substack{q_1, \dots, q_s \in \mathbb{Z}_+ \\ q_1 + \dots + q_s = p}} \binom{p}{q_1, \dots, q_s} \prod_{\ell=1}^s \mathbb{E} [(t_\ell u_{i,\ell})^{q_\ell}] \\ &= \sum_{\substack{q_1, \dots, q_s \in \mathbb{Z}_+ \\ q_1 + \dots + q_s = p}} \binom{p}{q_1, \dots, q_s} \prod_{\ell=1}^s t_\ell^{q_\ell} \left(\frac{q_\ell!}{a^{q_\ell r_\ell}} + o\left(\frac{1}{a^{q_\ell r_\ell}}\right) \right) = \frac{t^p p!}{a^{pr}} + o\left(\frac{1}{a^{pr}}\right), \end{aligned}$$

where $r = \min(r_1, \dots, r_s)$. Taking $k_p = pr$ and $C_p = t^p p!$ in Assumption 4.1 we see the condition $i \geq j, jk_i \geq ik_j$ is always verified. \square

Remark 4.13. Let us return to Corollary 4.7. This is a typical case and we take a close look of the density $f_a(x) = a(1 - x)^{a-1}$, $0 \leq x \leq 1$, of the Beta distribution $\text{Beta}(1, a)$.

For any continuous function $g : \mathbb{R} \rightarrow \mathbb{R}$, it is easy to verify that

$$\int_{\mathbb{R}} [g(x) - g(1)] f_a(x) dx = \int_0^1 [g(x) - g(1)] f_a(x) dx \rightarrow 0 \quad \text{as } a \rightarrow \infty.$$

This means that $\int_{\mathbb{R}} g(x) f_a(x) dx \rightarrow g(1)$. In other word, f_a converges to the Dirac delta function $\delta(x - 1)$. This observation hints that when the distribution f_a of v_i 's converges to the Dirac delta function $\delta(x - 1)$, or the random variable v_i converges in distribution to 1 (as $a \rightarrow \infty$) we should have the convergence of the random process $Q_{H,a}$. But we still need to impose some more technical conditions. We give a further illustration by the following corollary.

Corollary 4.14. Let the stick-breaking process P be defined as in Definition 2.1, where the corresponding v_i follows the following distribution:

$$f_a(x) = \begin{cases} a(1 - g(a)) & \text{if } 0 < x \leq 1/a; \\ \frac{ag(a)}{a-1} & \text{if } 1/a < x \leq 1, \end{cases}$$

where $g(a) = e^{-a^\epsilon}$, $a > 1$, for a certain arbitrarily fixed $\epsilon > 0$. Then, as $a \rightarrow \infty$, the conditions (4.3) and (4.4) of Theorem 4.4 hold for this density f_a . Thus the statements (4.5) and (4.7) of Theorem 4.4 hold true.

Proof. Before we proceed to the proof. Let us note the obvious fact that f_a converges to the Dirac delta distribution $\delta(x - 1)$.

For any $n > 0$, we see $\lim_{a \rightarrow \infty} a^n g(a) = 0$. A trivial calculation implies that for any positive integer p ,

$$\mathbb{E}[v_i^p] = \frac{(1/a)^p + g(a) \sum_{i=0}^p (1/a)^i}{p + 1} = \frac{1}{(p + 1)a^p} + o(b^{-p}).$$

An application of Assumption 4.1 with $k_p = p$ yields the desired statement. □

When the stick-breaking weights are i.i.d., Theorem 4.4 that we obtained for the stick-breaking random measure P covers very general situation and the conditions (4.3)–(4.4) are minimal and are easy to verify. But when the stick-breaking weights are not i.i.d. the situation becomes much more sophisticated like in other statistical situations. We shall consider some well-known processes introduced in the introduction. For these processes the explicit forms of the joint finite dimensional distributions of the stick-breaking weights, although complicated, are given (in the supplementary material). We can state similar results as those in Theorem 4.4 in one theorem for all these processes.

Theorem 4.15. Let P be one of the Poisson-Dirichlet process $\text{PDP}(a, b, H)$, the normalized inverse Gaussian process $\text{N-IG}(a, H)$, the normalized generalized gamma process $\text{NGG}(\sigma, a, H)$, and the generalized Dirichlet process $\text{GDP}(a, r, H)$. Then, we have the following results.

(i) As $a \rightarrow \infty$,

$$(D_a(A_1), D_a(A_2), \dots, D_a(A_n)) \xrightarrow{d} (X_1, X_2, \dots, X_n), \tag{4.10}$$

where $(X_1, X_2, \dots, X_n) \sim N(0, \Sigma)$ and $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq n}$ is given by

$$\sigma_{ij} = \begin{cases} 1 & \text{if } i = j \\ -\sqrt{\frac{H(A_i)H(A_j)}{(1-H(A_i))(1-H(A_j))}} & \text{if } i \neq j. \end{cases} \tag{4.11}$$

(ii) Let $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be the d -dimensional Euclidean space with the Borel σ -algebra. Then

$$Q_{H,a} \xrightarrow{\text{weakly}} B_H^o \quad \text{in} \quad D(\mathbb{R}^d) \tag{4.12}$$

with respect to the Skorohod topology.

Corollary 4.16. *Theorem 4.15 holds true when the random measure P is the Beta process (denoted by $P \sim \text{BP}(a, \gamma H)$), whose stick-breaking representation is given in Definition 7.8 (in the supplementary material). Our method of moments still works and in fact, due to the independence of the weights $w_{i,j}$ in (7.9) (in the supplementary material) the computation is much simpler.*

As long as the central limit theorem of P is obtained, it is trivial to use the delta-method to show the similar theorem for the nonlinear functional of this process. Using Theorem 3.9.4 in van der Vaart and Wellner (1996), we can state the following theorem.

Theorem 4.17. *Let P be one of $\text{N-IG}(a, H)$, $\text{PDP}(a, b, H)$, $\text{NGG}(\sigma, a, H)$, $\text{GDP}(a, r, H)$ or $\text{SPG}(\mu_a, H)$ satisfying (4.3)–(4.4) of Theorem 4.4. Let \mathbb{D} be the metric space of all probability measures on $(\mathbb{X}, \mathcal{X})$ with the total variation distance. Let $\phi : \mathbb{D} \rightarrow \mathbb{R}^d$ be a continuous functional which is Hadamard differentiable on \mathbb{D} . Then, as $a \rightarrow \infty$, we have*

$$\frac{1}{\sqrt{\mathbb{E}[\sum_{i=1}^{\infty} w_i^2]}} (\phi(P(\cdot)) - \phi(H(\cdot))) \xrightarrow{\text{weakly}} \phi'_{H(\cdot)}(B_H^o).$$

Remark 4.18. *When P is Dirichlet process or the normalized inverse Gaussian process, the above conclusion has been known (e.g. Labadi and Zarepour, 2013; Labadi and Abdelrazeq, 2016).*

One application of the above theorem is the limiting distribution of the quantile process of P .

Example 4.19. *Suppose $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and suppose that P is one of $\text{N-IG}(a, H)$, $\text{PDP}(a, b, H)$, $\text{NGG}(\sigma, a, H)$, $\text{GDP}(a, r, H)$ or $\text{SPG}(\mu_a, H)$ satisfying (4.3)–(4.4) of Theorem 4.4. Let H be absolutely continuous with positive derivative h . By Lemma 3.9.23 of van der Vaart and Wellner (1996), we have*

$$\frac{1}{\sqrt{\mathbb{E}[\sum_{i=1}^{\infty} w_i^2]}} (P^{-1}(\cdot) - H^{-1}(\cdot)) \xrightarrow{\text{weakly}} -\frac{B^o(\cdot)}{h(H^{-1}(\cdot))} = G(\cdot), \tag{4.13}$$

where $H^{-1}(s) = \inf\{t : H(t) \geq s\}$. The limiting process G is a Gaussian process with zero-mean and with covariance function

$$\text{Cov}(G((0, s]), G((0, t])) = \frac{s \wedge t - st}{h(H^{-1}((0, s])) h(H^{-1}((0, t]))}$$

for $s, t \in \mathbb{R}$,

5 Numerical illustration

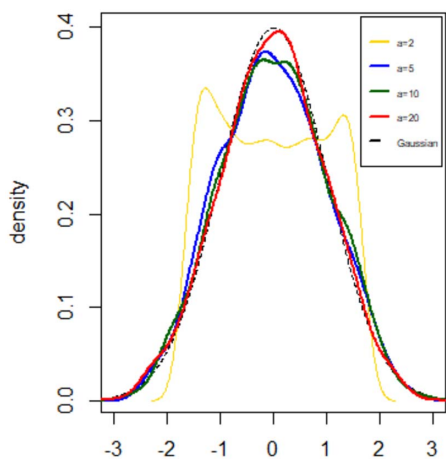
Theorem 4.15 states that $(D_a(A_1), D_a(A_2), \dots, D_a(A_n))$ converges to a joint normal distribution as $a \rightarrow \infty$. In this section we shall perform some numerical simulations to illustrate this convergence. To be specific the processes we choose to simulate are Poisson-Dirichlet process $\text{PDP}(a, b, H)$ and the stick-breaking process with general stick-breaking weights constructed in Corollary 4.14. For $\text{PDP}(a, b, H)$ we consider the cases the parameter $b = 0.2$ and $b = 0.5$ and for the process constructed in Corollary 4.14 we consider the cases $\epsilon = 1$ and $\epsilon = 5$.

For both of these two processes, the base measure H is assumed to be uniform distribution on $\mathbb{X} = (0, 1)$ and we take $n = 3$ and fix the partition of \mathbb{X} as $A_1 = (0, 0.3]$, $A_2 = (0.3, 0.7]$, $A_3 = (0.7, 1)$. In our simulations we truncate the infinite series (2.1) to 5000 terms and we simulate 2000 samples of $(D_a(A_1), D_a(A_2), D_a(A_3))$. Since it is rather messy to visualize the joint densities of $(D_a(A_1), D_a(A_2), D_a(A_3))$, we plot the histograms of the linear combination $1.6 \times D_a(A_1) + 1.4 \times D_a(A_2) + 0.5 \times D_a(A_3)$ (other linear combinations will produce similar results with different variances). The histograms for $\text{PDP}(a, b, H)$ with $b = 0.2$ and $b = 0.5$ and with $a = 2, 5, 10, 20$ are plotted in Figure 1 and the histograms for the stick-breaking process with general stick-breaking weights constructed in Corollary 4.14 with $\epsilon = 1$ and $\epsilon = 5$ and with $a = 2, 5, 10, 20$ are plotted in Figure 2. Graphs corresponding to different parameters (different b or different ϵ) are plotted in different figures but those corresponding to different values of a are plotted in the same figure with different colored curves so that one can observe the convergence to mean zero normal curves easily.

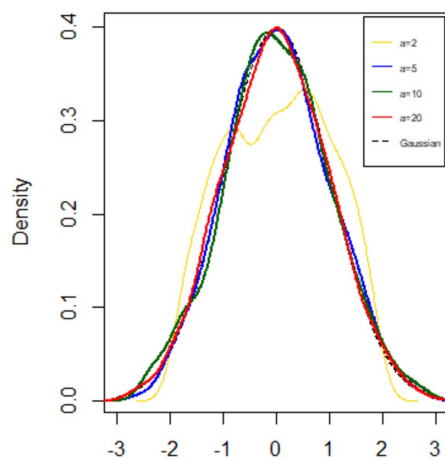
It is easy to observe that the convergence to the normal shape is very fast as a is getting larger.

6 Concluding remarks

The method of moments used in this paper could be applied to the study of asymptotics for some Bayesian nonparametric posterior processes in the following situations: (i) when the parameter a is finite and the sample size is large; (ii) when the parameter a is large and the sample size is finite; (iii) when the parameter a and the sample size are both large. As is well-known it is usually very hard to obtain the explicit form of the posterior distribution (even in the parametric cases) and even when the posterior distribution is obtained sometimes it is still very hard to use it to compute the needed statistics. A particularly interesting example is the posterior distribution of a homogeneous normalized random measure with independent increments (hNRMI) obtained by

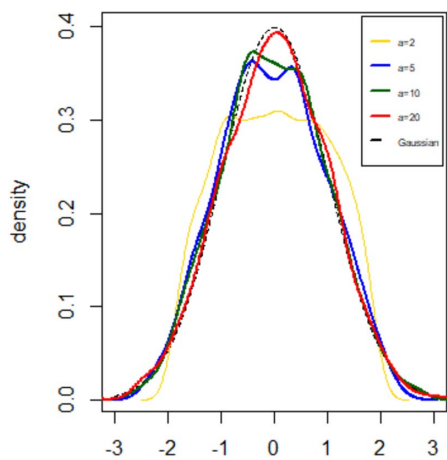


(a) $PDP(a, b, H)$ with $b = 0.2$.

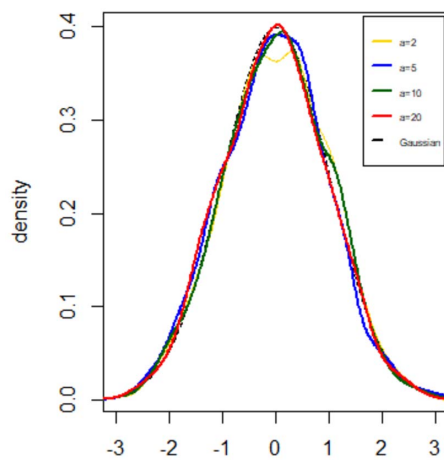


(b) $PDP(a, b, H)$ with $b = 0.5$.

Figure 1: Convergence of D_a with respect to $PDP(a, b, H)$ for $a = 2, 5, 10, 20$.



(a) Parameter $\epsilon = 1$.



(b) Parameter $\epsilon = 5$.

Figure 2: Convergence of D_a with respect to the constructed process in Corollary 4.14 for $a = 2, 5, 10, 20$.

James et al. (2009) and Favaro et al. (2016, Proposition 4). The hNRMI is a large class of priors, which contains the normalized generalized gamma processes (Definition 7.6 in the supplementary material) and the generalized Dirichlet processes (Definition 7.7 in the supplementary material) mentioned in this paper as special cases. Assume that P^a is an hNRMI depending on a parameter $a > 0$ and some other parameters studied in Favaro et al. (2016), where the parameter a is the same as the one in our paper when the hNRMI becomes a normalized generalized gamma process or a generalized Dirichlet process. If $\{X_i\}_{i=1}^n$ is a sample from the hNRMI P^a , in the sense that the sequence of exchangeable observations $\{X_i\}_{i=1}^n$ are defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in \mathbb{X} in such a way that, given P^a , X_1, \dots, X_n are i.i.d with distribution P^a , then the posterior distribution of P^a can be computed with the help of a latent variable U_n as follows (Favaro et al., 2016, Proposition 4)

$$P^a | U_n, X_1, \dots, X_n \sim \hat{P}_n^a := \varphi_{0, U_n}^a \tilde{P}_{U_n}^a + \sum_{j=1}^k \varphi_{j, U_n}^a \delta_{X_j^*},$$

where given $U_n = u$, \tilde{P}_u^a is an hNRMI admitting a stick-breaking representation and $\{X_j^*\}_{j=1}^k$ are the distinct values of $\{X_i\}_{i=1}^n$. To compute a Bayesian statistic, we need to compute some functional of the posterior probability measure \hat{P}_n^a . For example, to find the quantile t_q such that $\hat{P}_n^a((-\infty, t_q)) \leq q \leq \hat{P}_n^a((-\infty, t_q])$ for some given $q \in (0, 1)$ or to compute $\int_{\mathbb{X}^d} f(x_1, \dots, x_d) \hat{P}_n^a(dx_1) \cdots \hat{P}_n^a(dx_d)$ and so on (e.g. Ferguson, 1973; Hu and Zhang, 2020), which is usually complicated due to the complexity of $\tilde{P}_{U_n}^a$. However, when a is sufficiently large the probability measure $\tilde{P}_{U_n}^a$ is approximately a normal distribution, then we can use the normal distribution to approximate $\tilde{P}_{U_n}^a$ in the computation of these Bayesian statistics. Let us also point out that in some situations the normal approximation is sufficiently good for reasonable size a (from the figures in this paper, we see that when $a = 5$, the graphs are already close to normal distributions).

Supplementary Material

Supplementary Material of “Functional central limit theorems for stick-breaking priors” contains all proofs of the results provided in the main paper, together with the definitions and some well-known propositions of the studied processes. (DOI: [10.1214/21-BA1290SUPP](https://doi.org/10.1214/21-BA1290SUPP); .pdf).

References

- Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition. A Wiley-Interscience Publication. [MR1324786](#). 1103
- Brix, A. (1999). “Generalized Gamma measures and shot-noise Cox processes.” *Advances in Applied Probability*, 31(04): 929–953. [MR1747450](#). doi: <https://doi.org/10.1239/aap/1029955251>. 1103

- Brunner, L. J. and Lo, A. Y. (1996). “Limiting posterior distributions under mixture of conjugate priors.” *Statistica Sinica*, 6(1): 187–197. [MR1379056](#). 1102
- Dawson, D. A. and Feng, S. (2006). “Asymptotic behavior of the Poisson–Dirichlet distribution for large mutation rate.” *The Annals of Applied Probability*, 16(2): 562–582. [MR2244425](#). doi: <https://doi.org/10.1214/105051605000000818>. 1103
- De Blasi, P., Lijoi, A., and Pruenster, I. (2013). “An asymptotic analysis of a class of discrete nonparametric priors.” *Statistica Sinica*, 23(3): 1299–1322. [MR3114715](#). 1102
- Favaro, S., Lijoi, A., Nava, C., Nipoti, B., Prünster, I., and Teh, Y. W. (2016). “On the stick-breaking representation for homogeneous NRMIs.” *Bayesian Analysis*, 11(3): 697–724. [MR3498043](#). doi: <https://doi.org/10.1214/15-BA964>. 1102, 1117
- Favaro, S., Lijoi, A., and Prünster, I. (2012). “On the stick-breaking representation of normalized inverse Gaussian priors.” *Biometrika*, 99(3): 663–674. [MR2966776](#). doi: <https://doi.org/10.1093/biomet/ass023>. 1102
- Ferguson, T. S. (1973). “A Bayesian analysis of some nonparametric problems.” *The Annals of Statistics*, 1(2): 209–230. [MR0350949](#). 1101, 1103, 1117
- Ghosal, S. and Van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*, volume 44. Cambridge: Cambridge University Press, first edition. [MR3587782](#). doi: <https://doi.org/10.1017/9781139029834>. 1105
- Ho Jang, G., Lee, J., and Lee, S. (2010). “Posterior consistency of species sampling priors.” *Statistica Sinica*, 20(2): 581–593. [MR2682631](#). 1102
- Hu, Y. and Zhang, J. (2020). “Dirichlet process and Bayesian nonparametric models (in Chinese).” *Science in China, Chinese edition*, To appear: 1–41. 1105, 1117
- Hu, Y. and Zhang, J. (2021). “Supplementary material for: Functional Central Limit Theorems for Stick-Breaking Priors.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/21-BA1290SUPP>. 1105
- James, L. F. (2008). “Large sample asymptotics for the two-parameter Poisson–Dirichlet process.” In *Clarke, B. and Ghosal, S. (eds.), Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, volume 3, 187–199. Cambridge, MA: MIT Press. [MR2459225](#). doi: <https://doi.org/10.1214/074921708000000147>. 1102
- James, L. F., Lijoi, A., and Prünster, I. (2009). “Posterior analysis for normalized random measures with independent increments.” *Scandinavian Journal of Statistics*, 36(1): 76–97. [MR2508332](#). doi: <https://doi.org/10.1111/j.1467-9469.2008.00609.x>. 1117
- Kim, Y. and Lee, J. (2004). “A Bernstein–von Mises theorem in the nonparametric right-censoring model.” *The Annals of Statistics*, 32(4): 1492–1512. [MR2089131](#). doi: <https://doi.org/10.1214/009053604000000526>. 1102
- Labadi, L. A. and Abdelrazeq, I. (2016). “On functional central limit theorems of

- Bayesian nonparametric priors." *Statistical Methods & Applications*, 26(2): 215–229. MR3652494. doi: <https://doi.org/10.1007/s10260-016-0365-8>. 1103, 1114
- Labadi, L. A. and Zarepour, M. (2013). "On asymptotic properties and almost sure approximation of the normalized inverse-Gaussian process." *Bayesian Analysis*, 8(3): 553–568. MR3102225. doi: <https://doi.org/10.1214/13-BA821>. 1103, 1114
- Lijoi, A., Mena, R. H., and Prünster, I. (2005a). "Bayesian nonparametric analysis for a generalized Dirichlet process prior." *Statistical Inference for Stochastic Processes*, 8(3): 283–309. MR2177315. doi: <https://doi.org/10.1007/s11203-005-6071-z>. 1103
- Lijoi, A., Mena, R. H., and Prünster, I. (2005b). "Hierarchical mixture modeling with normalized inverse-Gaussian priors." *Journal of the American Statistical Association*, 100(472): 1278–1291. 1103
- Lijoi, A., Mena, R. H., and Prünster, I. (2007). "Controlling the reinforcement in Bayesian non-parametric mixture models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4): 715–740. MR2370077. doi: <https://doi.org/10.1111/j.1467-9868.2007.00609.x>. 1103
- Lijoi, A., Prünster, I., et al. (2003). "On a normalized random measure with independent increments relevant to Bayesian nonparametric inference." In *13th European Young Statisticians Meeting*, 123–134. Staempfli. 1103
- Lo, A. Y. (1983). "Weak convergence for Dirichlet processes." *Sankhyā: The Indian Journal of Statistics, Series A*, 105–111. MR0749358. 1102
- Lo, A. Y. (1987). "A large sample study of the Bayesian bootstrap." *The Annals of Statistics*, 15(1): 360–375. MR0885742. doi: <https://doi.org/10.1214/aos/1176350271>. 1102
- Patrick, B. (1995). *Probability and Measure*. New York: John Wiley & Sons, third edition. 1109
- Perman, M., Pitman, J., and Yor, M. (1992). "Size-biased sampling of Poisson point processes and excursions." *Probability Theory and Related Fields*, 92(1): 21–39. MR1156448. doi: <https://doi.org/10.1007/BF01205234>. 1102
- Pitman, J. (1996). "Some developments of the Blackwell-MacQueen urn scheme." In *Statistics, Probability and Game Theory, Lecture Notes-Monograph Series*, volume 30, 245–267. Hayward, CA. MR1481784. doi: <https://doi.org/10.1214/lnms/1215453576>. 1109
- Pitman, J. (2003). "Poisson-Kingman partitions." In *Statistics and Science: a Festschrift for Terry Speed, Lecture Notes-Monograph Series*, volume 40, 1–34. Beachwood, OH. MR2004330. doi: <https://doi.org/10.1214/lnms/1215091133>. 1109
- Pitman, J. and Yor, M. (1997). "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator." *The Annals of Probability*, 25(2): 855–900. MR1434129. doi: <https://doi.org/10.1214/aop/1024404422>. 1103

- Regazzini, E., Lijoi, A., and Prünster, I. (2003). “Distributional results for means of normalized random measures with independent increments.” *The Annals of Statistics*, 31(2): 560–585. MR1983542. doi: <https://doi.org/10.1214/aos/1051027881>. 1102
- Sethuraman, J. (1994). “A constructive definition of Dirichlet priors.” *Statistica Sinica*, 639–650. MR1309433. 1102
- Sethuraman, J. and Tiwari, R. C. (1982). “Convergence of Dirichlet measures and the interpretation of their parameter.” In *Statistical Decision Theory and Related Topics III*, 305–315. Elsevier. MR0705321. 1102
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer. MR1385671. doi: <https://doi.org/10.1007/978-1-4757-2545-2>. 1114
- Zeng, C. and Duan, L. L. (2020). “Quasi-Bernoulli Stick-breaking: Infinite Mixture with Cluster Consistency.” *arXiv preprint arXiv:2008.09938*. 1106

Acknowledgments

We sincerely thank the anonymous referees for the careful reading and for the constructive and inspiring comments which improves the paper significantly.